**Motivation**

The betweenness centrality of a node \( u \) is defined as

\[
B(u) = \sum_{s,t} \frac{\sigma_{s,t}(u)}{\sigma_{s,t}},
\]

where \( \sigma_{s,t} \) is the number of \( s-t \) shortest paths, and \( \sigma_{s,t}(u) \) is the number of \( s-t \) shortest paths that have \( u \) as their internal node.

- **Community detection**: Betweenness centrality is frequently used to detect communities in large scale networks [3].

- **Navigation applications**: It is also used as a successful heuristic for selecting landmarks in state-of-the-art shortest path applications [1].

- **Attacking graph connectivity**: Real-world networks are robust to random failures but fragile with respect to targeted attacks. Betweenness centrality is used as a good heuristic to destroy connectivity.

**Main contributions**

For \( S \subseteq V \), we define the betweenness centrality of \( S \) as

\[
B(S) = \sum_{s,t} \frac{\sigma_{s,t}(S)}{\sigma_{s,t}},
\]

where \( \sigma_{s,t}(S) \) is the number of \( s-t \) shortest paths that have an internal node in \( S \).

**Contribution 1**

Prior work on BWC estimation strongly relies on the assumption that \( \text{OPT}_k = \Theta(n^2) \) for a constant integer \( k \) [4]. We show this assumption is not true in general.

We explain empirical evidence which supports this strong assumption using Random Apollonian Networks that provably generate scale-free, small-world graphs with high probability [2]. Also, bounded-tree width networks including Barabasi-Albert random graphs satisfy this assumption.

**Contribution 2**

We design \textsc{Hedge} – a \((1 - 1/e - \epsilon)\)-approximation algorithm – that uses smaller sized samples compared to state-of-the-art [4].

Our proposed method outperforms the state-of-the-art method due to Yoshida [4].

**Contribution 3**

We provide a general analytical framework based on Chernoff bound and submodular optimization, and show that it can be applied to any other centrality measure if it is monotone-submodular, and (ii) admits a hyper-edge sampler.

The size of the largest connected component, as we remove the first 1000 nodes in the order induced by centralities.

**Experimental Results**

**HEDGE vs. EXHAUST** (baseline method): centralities and speedups.

**Experimental results**

- **Time evolving networks**:
  - (a) AS: \( k = 1 \)
  - (b) AS: \( k = 50 \)
  - (c) DBLP: \( k = 1 \)
  - (d) DBLP: \( k = 50 \)
  - (e) KG: \( k = 1 \)
  - (f) KG: \( k = 50 \)

Largest betweenness centrality score and number of nodes, edges and average degree versus time on the (i) Autonomous systems (a),(b) (ii) DBLP dataset (c),(d) and (iii) stochastic Kronecker graphs (e),(f).

**References**


